

⇒[MetaMap Home](#)
⇒[FAQ Home](#)
⇒[Help/Questions](#)

MetaMap's Base, USAbase, and NLM Data Versions

In order to accommodate the UMLS source-vocabulary licensing permissions and processing requirements of as many users as possible, MetaMap produces three distinct versions of UMLS data that are based on the Restriction Categories of Metathesaurus source vocabularies. (See the Restriction Categories tab on the [UMLS Source Release Documentation Page](#) for all UMLS sources' Restriction Categories.) Each data version includes both [Strict and Relaxed Models](#).

Listed from smallest to largest, the three data versions are:

1. **Base:** The Base data version includes those source vocabularies with no associated licensing restrictions beyond those of the [UMLS Metathesaurus License](#)—all and only sources of Restriction Category 0.
2. **USAbase:** The USAbase data version includes those source vocabularies with no associated restrictions beyond a [UMLS Metathesaurus License](#), and free for use for US-based projects; this version includes the Base vocabularies (those with Restriction Category 0), plus the five Category-4 sources and the four Category-9 sources (including, most notably, SNOMEDCT). The USAbase version is a proper superset of the Base version, and might be the most appropriate version for users with a [SNOMED CT License](#). This data version is MetaMap's default, but the default can be overridden by specifying another user-installed data version using the `-V` flag, as described below.
Users without SNOMED-CT licenses should use Base data version.
3. **NLM:** Beginning with the 2015AA UMLS release, the NLM data version includes the full Metathesaurus. Previous versions of the NLM model excluded the CPT, CPTSP, HCPT, and MTHCH vocabularies from the CPT family, and the HCDT, HCPCS, and MTHHH vocabularies from the HCPCS family.

MetaMap users are expected to adhere to all source-vocabulary licenses.

As noted above, the USAbase data version is MetaMap's default. To run MetaMap with the Base data version, simply call `metamap -V Base`; similarly, to run MetaMap with the NLM data version, simply call `metamap -V NLM`.

Users with the appropriate licenses may find it useful to restrict MetaMap's output to certain UMLS source vocabularies using the `--restrict-to.sources` command-line option in conjunction with the `-V` flag.

For example, to map text to only concepts in the MedDRA source vocabulary, MetaMap should be called as `metamap -J MDR -V NLM`

because MedDRA has Restriction Category 3, and is therefore in only the NLM Data Version.

All MetaMap data versions (Base, USAbase, and NLM) include only the English strings from the UMLS sources (i.e., `LAT == "ENG"`). Users needing to include sources in languages other than English in order to process non-English text should create their own UMLS-like database using our Data File Builder suite, available [here](#).